



# INU3011 Documents structurés

## Plan de cours Hiver 2014

[Yves MARCOUX](#) - [EBSI](#) - [Université de Montréal](#)

---

## Informations générales

- Cours de trois crédits, ce qui correspond à 135 heures de travail pour l'étudiant.
  - [Description officielle](#) à l'[Annuaire général de l'Université](#):  

Formats de documents et langages de balisage. Historique. Concepts de base. XML et normes périphériques. Modélisation, validation, stylage. Chaînes de traitement et méthodologies d'implantation. Bases de données XML. Préalable : INU1001 ou l'équivalent.
  - **Horaire et locaux:** Lundi de 16h à 18h50 au local B3245 du [Pavillon 3200, rue Jean-Brillant](#). Certains travaux de laboratoire s'effectuent, à l'intérieur de cette même plage horaire, au local C2031 du [Pavillon Lionel-Groulx](#) (un des [Laboratoires d'informatique documentaire de l'EBSI](#); entrée par le C2035). *À moins d'avis contraire, le cours débute toujours au B3245 et le groupe se déplace au C2031 lorsque nécessaire.*
  - **Professeur:** [Yves MARCOUX](#) <[y marcoux@gmail.com](mailto:y marcoux@gmail.com)>.
  - **Site du cours:** <<http://cours.ebsi.umontreal.ca/INU3011/>>.
- 

## Objectifs

À la fin du cours, l'étudiant(e) démontrera une compréhension approfondie des enjeux liés à l'utilisation des documents structurés pour la gestion de l'information documentaire. Plus spécifiquement, l'étudiant(e):

- Démontrera une maîtrise approfondie des concepts et des principes de base des documents structurés.
- Pourra concevoir des modèles de documents structurés en fonction de besoins donnés.
- Pourra créer des feuilles de styles en fonction de besoins donnés.
- Pourra planifier des chaînes de traitement de documents structurés en fonction de besoins donnés.
- Démontrera une bonne compréhension du processus d'implantation d'un système de documents

structurés.

---

## Description

Tel qu'indiqué ci-dessus, l'objectif principal de ce cours est de faire prendre conscience aux participant(e)s des avantages fondamentaux liés à l'utilisation des documents structurés pour la publication d'information documentaire, ainsi que des défis qu'ils posent aux concepteurs de systèmes d'information. Un objectif secondaire est la familiarisation avec certains types de fonctionnalité des outils commerciaux de traitement des documents structurés.

Le cours est presque entièrement élaboré autour de XML (*Extensible Markup Language*), le plus important format normalisé de documents structurés existant actuellement. Outre une compréhension approfondie des concepts et principes de base des documents structurés, on vise un niveau de compétence permettant de concevoir des DTD (*Document Type Definitions*) et/ou des schémas XML (du W3C), de créer des feuilles de styles CSS et XSLT, de sélectionner des outils en fonction de besoins spécifiques, de convertir des documents XML en HTML et de publier des documents XML sur le web, directement et par le truchement d'une base de données XML.

Une grande autonomie et un haut niveau d'implication et de participation (en classe **et surtout au laboratoire**) sont attendus des étudiant(e)s.

*Une connaissance de XHTML et de CSS équivalente aux éléments de ces langages vus dans le cours INU1001 Introduction à l'information numérique (ou un équivalent, comme SCI6052) est un prérequis essentiel, y compris pour les quiz en classe. Veuillez au besoin réviser ces notions **avant le Cours 2.***

L'activité principale est un projet (réalisé en équipes de deux) consistant à modéliser des documents, à les saisir en XML et à les publier (sur le web ou sur un autre support). La nature des documents à traiter sera discutée en classe.

Il pourrait y avoir un conférencier invité à parler d'une utilisation concrète des documents structurés (l'horaire serait alors réaménagé en conséquence).

---

## Calendrier des activités

### Notes:

- Certaines parties du cours se déroulent au laboratoire C2031 (voir [Horaire et locaux](#) ci-dessus).
- Les contenus, incluant le travail supervisé en laboratoire, et leur répartition entre les cours seront constamment ajustés au cours de la session.

### Calendrier des activités

Date – Cours	Contenu et lectures	Évaluations
2014-01-13 – C1	Introduction au cours et à XML. Contextes et historique de XML. Notions syntaxiques de base.	
2014-01-20 – C2	Préparation au Quiz 1. Morphologie d'un document XML, encodage. Balisage, contenu, contenu textuel. Documents XML comme structures hiérarchiques.  <b>Lectures obligatoires:</b> <a href="#">XML Essentials</a> , <a href="#">XML en 10 points</a> , <a href="#">Morphologie et encodage d'un document XML</a> , <a href="#">Introduction aux entités en XML</a> , <a href="#">XML et les navigateurs web Firefox et Internet Explorer</a>	
2014-01-27 – C3	Quiz 1 (en classe).  Entités déclarées, internes et externes.	Quiz 1  Remise (sur votre espace web gin-ebis + courriel au professeur) du travail de <a href="#">Validation</a>

		<a href="#">des prérequis</a>
2014-02-03 – C4	<p>Applications validantes et non-validantes, documents valides versus bien-formés, définitions de type de documents, schémas XML. Déclarations d'éléments et modèles de contenu. Déclarations d'attributs.</p> <p><b>Lecture obligatoire:</b> <a href="#">Bien-formé versus validité d'un document XML</a></p>	
2014-02-10 – C5	Exploitation de la validité en environnement auteur.	
2014-02-17 – C6	<p>Processus et principes de modélisation en XML. Préparation au Quiz 2.</p> <p><b>Lecture obligatoire:</b> <a href="#">Attributs et modélisation en XML</a></p>	
2014-02-24 – C7	<p>Quiz 2 (en classe).</p> <p>Modélisation: quelques points spécifiques. Contenus multimédias, liens entre documents. Documents à cycle de vie. Modèles XML normalisés: la <i>Text Encoding Initiative</i> (TEI).</p> <p><b>Lecture obligatoire:</b> <a href="#">ROMARY - Questions &amp; Answers for TEI Newcomers</a></p>	Quiz 2
2014-03-03	<i>Semaine de lecture — Pas de cours</i>	
2014-03-10 – C8	Traitement de documents XML: validation et stylage, en navigateur et avec outils spécialisés. XPath et XSLT (début).	
2014-03-17 – C9	<p>XPath et XSLT (fin, en classe). Aperçu de LaTeX, configuration d'oXygen pour produire du LaTeX.</p> <p>Modélisation et stylage (en laboratoire).</p> <p><b>Lectures obligatoires:</b> <a href="#">LaTeX Project - An introduction to LaTeX</a>, <a href="#">UNWALLA - LaTeX: an introduction</a></p>	
2014-03-24 – C10	Modélisation et stylage (en laboratoire).	Remise (sur <a href="#">StudiUM</a> ) du rapport préliminaire de modélisation
2014-03-31 – C11	Quelques points sur XSLT.	
2014-04-07 – C12	Publication de documents XML sur le web. Traitements en lot. Bases de données XML.	

	Modélisation et stylage (en laboratoire).	
2014-04-14 – C13	Évaluation du cours. Publication web (en laboratoire).	
2014-04-21	Lundi de Pâques – pas de cours	
2014-04-28 – C14	Chaînes types de traitement de documents XML. Méthodes de création de documents XML: exportation, conversion, « upconversion ». Méthodologies de mise sur pied de systèmes de documents structurés. Modèles XML normalisés (autres que TEI). RDF et web sémantique (si le temps le permet).	
2014-05-03 (samedi)	Dernier jour de la session	Remise (sur <a href="#">StudiUM</a> ) du rapport de modélisation et stylage  Remise (sur <a href="#">StudiUM</a> ) du rapport de publication web

## Évaluation

### Évaluation

Activité	Pondération
Travail individuel de validation des prérequis	5%
Deux quiz individuels à livres fermés en classe Quiz 1: 15%, Quiz 2: 15%	30%
Rapport préliminaire de modélisation	10%
Travail de modélisation et stylage	35%
Travail de publication web	20%

Les activités évaluées se font en équipe de deux, sauf le travail de validation des prérequis et les quiz. Les travaux de validation des prérequis, de modélisation et stylage et de publication web sont décrits dans des protocoles rendus disponibles en temps opportun.

L'évaluation avec seuil (voir [section à cet effet](#) ci-après) est appliquée dans ce cours.

La méthode de calcul de la note finale du cours est présentée au

## Règlements, politiques

### Évaluation avec seuil

L'évaluation avec seuil, telle que définie à l'EBSI (voir par exemple le [Guide de l'étudiant du certificat en GIN](#)), est appliquée dans ce cours, c'est-à-dire que pour réussir le cours, il faut non seulement atteindre (ou dépasser) la note de passage sur l'ensemble des évaluations du cours, mais également atteindre (ou dépasser) cette même note de passage sur le *sous-ensemble (pondéré) des activités évaluées qui sont réalisées individuellement (et non en équipe)*.

Notez que la note de passage est de « D » pour les étudiants inscrits à un programme de premier cycle et de « C » pour les étudiants inscrits à un programme des cycles supérieurs.

### Remise des travaux à évaluer

Les artefacts à évaluer (fichiers, rapports, etc.) doivent être remis au professeur, ou déposé sur [StudiUM](#) **en un seul fichier** (compressé si nécessaire), *avant le début du cours* (s'il y a cours; autrement, à midi) le jour prévu pour la remise. Ils doivent être remis **en format numérique**, conformément aux directives indiquées dans le protocole ou l'énoncé du travail concerné.

Si la remise comporte un rapport (le cas échéant, cela est spécifié dans le protocole ou l'énoncé du travail), ce rapport doit être en format PDF, OpenOffice ou Word, prêt à imprimer sur papier 8,5 x 11 po. Les rapports doivent impérativement être structurés en sections numérotées et paginés.

Il est également possible, mais seulement après entente avec le professeur, de remettre des artefacts à évaluer en format papier.

**Si vous ne respectez pas les directives ci-dessus, vous êtes considéré comme n'ayant pas remis le travail.**

Vous devez *impérativement* conserver une copie de sécurité intégrale des artefacts remis, *au moins jusqu'à réception de leur correction*. Vous devez être prêt à transmettre rapidement sur demande au correcteur (par courriel ou autrement) une copie du matériel remis originellement, en cas de problème de lecture ou autre; c'est votre responsabilité. Pour les artefacts papier, il est **très fortement recommandé** d'en conserver une copie de sécurité numérique (fichier) complète ou une photocopie, de façon à pouvoir en produire facilement un nouvel exemplaire en cas de problème.

### Mode de communication

Le mode de communication privilégié du professeur vers les étudiant(e)s est la section [Nouvelles concernant le cours](#) de la [page d'accueil du cours](#). Vous pouvez, si désiré, vous abonner à ces nouvelles

par le truchement du [fil RSS associé](#); sinon, vous devez consulter la section *Nouvelles* très régulièrement (au moins une fois par jour).

Le professeur peut également utiliser *le courriel* pour joindre les étudiant(e)s. L'adresse utilisée pour vous joindre sera celle enregistrée dans votre profil informatique à la DGTIC <[http://www.dgtic.umontreal.ca/profil\\_usager/profil\\_dgtic.htm](http://www.dgtic.umontreal.ca/profil_usager/profil_dgtic.htm)>; assurez-vous que cette adresse soit valide et fonctionnelle en tout temps (voyez notamment à ne pas laisser votre boîte aux lettres se remplir). Vous devez lire très régulièrement votre courriel (au moins une fois par jour).

Le mode de communication privilégié des étudiant(e)s vers le professeur est le courriel. L'adresse courriel du professeur est <[ymarcoux@gmail.com](mailto:y Marcoux@gmail.com)>. **SVP, toujours inscrire la mention [INU3011] (incluant les crochets) au début de la ligne sujet de votre message.**

## **Autres règlements et politiques**

**Tous les règlements, politiques et directives énoncés dans le [Guide de l'étudiant du certificat en GIN](#) s'appliquent.** Une attention particulière est à porter aux éléments suivants :

### *Travaux en équipe*

Pour les travaux réalisés en équipe, le professeur se réserve le droit d'évaluer séparément chaque membre d'une équipe.

### *Code d'honneur*

Il est attendu que les étudiant(e)s respectent le [code d'honneur de l'EBSI](#).

## Règlement disciplinaire sur le plagiat ou la fraude concernant les étudiants

Toute infraction au règlement sur le plagiat ou la fraude sera traitée suivant la procédure indiquée dans le règlement.

## Retard dans la remise des travaux

Tout retard non justifié dans la remise d'un travail sera sanctionné : 5% de la note est retranché par jour de calendrier de retard jusqu'à un maximum de 35%; à la 8<sup>ième</sup> journée de calendrier, la note F ou zéro (0) est attribuée.

## Qualité de la langue

Un maximum de 10% de la note globale d'un travail pourra être retranché pour mauvaise qualité de la langue dans les travaux (ne s'applique pas aux examens ou quiz en classe).

---

# Bibliographie

## Lectures obligatoires (par ordre de lecture dans le cours)

- W3C. *XML Essentials*. En ligne <<http://www.w3.org/standards/xml/core>>.
- W3C. *XML en 10 points*. En ligne <<http://www.w3.org/XML/1999/XML-in-10-points.fr.html>>.
- MARCOUX, Yves. « Morphologie et encodage d'un document XML ». En ligne <<http://cours.ebsi.umontreal.ca/INU3011/H2014/matthem/morpho.htm>>.
- MARCOUX, Yves. « Introduction aux entités en XML ». En ligne <<http://cours.ebsi.umontreal.ca/INU3011/H2014/matthem/entites.htm>>.
- MARCOUX, Yves. « XML et les navigateurs web Firefox et Internet Explorer ». En ligne <<http://cours.ebsi.umontreal.ca/INU3011/H2014/divers/outils-XML-IE.htm>>.
- MARCOUX, Yves. « Bien-formé versus validité d'un document XML ». En ligne <<http://cours.ebsi.umontreal.ca/INU3011/H2014/matthem/bien-forme-vs-validite.htm>>.
- MARCOUX, Yves. « Attributs et modélisation en XML ». En ligne <<http://cours.ebsi.umontreal.ca/INU3011/H2014/matthem/attrib-modelisation.htm>>.
- ROMARY, Laurent. "Questions & Answers for TEI Newcomers." *Jahrbuch für Computerphilologie 10*. Mentis Verlag, 2009. En ligne <<http://arxiv.org/abs/0812.3563>>.
- LaTeX Project. "An introduction to LaTeX." *LaTeX project site*, février 2008. En ligne <<http://latex-project.org/intro.html>>.
- UNWALLA, Mike. "LaTeX: an introduction." *Communicator*, printemps 2006. En ligne <<http://www.techscribe.co.uk/ta/latex-introduction.pdf>>.

## Lectures recommandées

### Notes:

- Les tutoriels de W3Schools sont particulièrement recommandés.
- Plusieurs autres ressources sur XSL/XSLT sont répertoriées dans [Atrium](#).
- BITOUZÉ, Denis; CHARPENTIER, Jean-Côme. *LaTeX, l'essentiel : pour une prise en main rapide et efficace*. Paris : Pearson education c2010. (Réserve Z 253.4 L38 B59 2010 DVDROM)

- GLUSHKO, Robert J.; McGRATH, Tim. *Document engineering: analyzing and designing documents for business informatics and Web services*. The MIT Press, 2005, 703 p. En ligne (de umontreal.ca) <<http://www.books24x7.com/marc.asp?isbn=0262072610>>. (Réserve QA 76.9 T48 G58 2005)
- KAY, Michael. *XSLT: programmer's reference (2nd edition)*. Wrox Press, 2001, 939 p. En ligne (de umontreal.ca) <<http://www.books24x7.com/marc.asp?isbn=0764543814>>. (Réserve QA 76.73 X58 K39 2001)
- KAY, Michael. *XSLT 2.0: programmer's reference (3rd edition)*. Wiley, 2004. En ligne (de umontreal.ca) <<http://www.books24x7.com/marc.asp?isbn=0764569090>>.
- MALER, Eve; EL ANDALOUSSI, Jeanne. *Developing SGML DTDs: From Text to Model to Markup*. Prentice Hall PTR, 1996. (Réserve QA 76.76 H94 M26 1996)
- MILLER, Dick R.; CLARKE, Kevin S. *Putting XML to Work in the Library: Tools for Improving Access and Management*. Chicago, ALA, 2004. (Réserve Z 678.93 X54 M55 2004)
- NELLHAUS, Tobin. "XML, TEI, Digital Libraries in the Humanities." *Portal: Libraries and the Academy* 1.3 (2001): 267-277. En ligne <[http://muse.jhu.edu/journals/portal\\_libraries\\_and\\_the\\_academy/v001/1.3nellhaus.html](http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v001/1.3nellhaus.html)>.
- RAY, Erik T. *Markup and Core Concepts*. En ligne <<http://oreilly.com/catalog/learnxml/chapter/ch02.html>>.
- SAMPSON, Geoffrey. *e.biz: The Anatomy of Electronic Business*. Elsevier, 2005. (Réserve HF 5548.32 S26 2004)
- Text Encoding Initiative Consortium. *A Gentle Introduction to XML*. En ligne <<http://xml.coverpages.org/TEI-GentleIntroXML.pdf>>.
- TENNANT, Roy. *XML in Libraries*. Neal-Schuman, 2002. (Réserve Z 678.93 X54 X65 2002)
- VAN DONGEN, M.R.C. *LaTeX and friends*. Heidelberg ; New York : Springer c2012, accessible en ligne via Atrium.
- W3Schools. "XML Tutorial." En ligne <<http://www.w3schools.com/xml/>>.
- W3Schools. "XPath Tutorial." En ligne <<http://www.w3schools.com/xpath/>>.
- W3Schools. "XSLT Tutorial." En ligne <<http://www.w3schools.com/xsl/>>.

## Lectures utiles

- DVORAK, John C. "Killing the Web," *PC Magazine*, octobre 2000. En ligne <<http://www.pcmag.com/article2/0,2817,31944,00.asp>>.
- GRDS. [XML en route au Gouvernement du Québec](#) (rapport du <GRDS> sur l'utilisation de XML au Gouvernement du Québec; format PDF)
- GRDS. [Cadre de référence gouvernemental en gestion intégrée des documents](#) (rapport du <GRDS> sur la gestion intégrée des documents au Gouvernement du Québec; format PDF)
- KARP, Alan H. "Viewpoint: Making money selling content that others are giving away." *Communications of the ACM*, jan. 2003. En ligne <<http://portal.acm.org/citation.cfm?doid=602421.602438>>.
- KORPELA, Jukka. *A tutorial on character code issues*. En ligne <<http://www.cs.tut.fi/~jkorpela/chars.html>>. Mises à jour régulières.
- MARCOUX, Yves. « Les formats normalisés de documents électroniques ». *ICO Québec*, vol. 6, nos 1-2, printemps 1994, pp. 56-65. En ligne <<http://marcoux.ebsi.umontreal.ca/publi/ico1994/ico94.htm>>.
- MARCOUX, Yves. « Place de SGML parmi les nouvelles architectures documentaires ». *Actes de la conférence « Technologie SGML 1996 » organisée par le Centre de recherche en droit public de l'Université de Montréal*, CRDP, 1996, pp. 1-13. En ligne <<http://marcoux.ebsi.umontreal.ca/publi/ottawa/marcoux.html>>.
- MARCOUX, Yves; SÉVIGNY, Martin. "Why SGML? Why now?" *Journal of the American Society for Information Science*, vol. 48, no 7, July 1997, Special Topic Issue: Structured Information/Standards for Document Architectures, pp. 584-592.
- NELSON, Theodor Holm. "Embedded Markup Considered Harmful." En ligne: <<http://www.xml.com/pub/a/w3j/s3.nelson.html>>.
- PILLOU, Jean-François. *Introduction à XML*. En ligne <<http://www.commentcamarche.net/xml/xmlintro.php3>>.
- ROUMIEUX, Olivier. « XML sans frontières ». *Archimag*, no 159, novembre 2002.
- RYMER, John R. "Why 90 percent of XML standards will fail." En ligne <<http://www.zdnet.com/news/why-90-percent-of-xml-standards-will-fail/114625>>.

## Diverses sources et ressources utiles

- ANDRIES, Patrick; et al. Une traduction en français de la norme XML 1.0. En ligne

- <<http://pages.videotron.com/fyergeau/w3c/xml10/REC-xml-19980210.fr.html>>.
- BRAY, Tim. La norme XML annotée par un des éditeurs, Tim Bray. En ligne <<http://www.xml.com/axml/axml.html>>.
  - COVER, Robin. “The Cover Pages: The Online Resource for Markup Language Technologies.” En ligne <<http://xml.coverpages.org/>>.
  - GRDS. « Groupe départemental de Recherche sur les Documents Structurés ». En ligne <<http://grds.ebsi.umontreal.ca/>>.
  - MICHARD, Alain. *XML: Langage et applications*. Nouvelle édition, Eyrolles, 2001, 499 p. (Réserve QA 76.76 H94 M53 2001)
  - SOLARI, J.J. Une traduction en français de la norme XML 1.1. En ligne <<http://www.yoyodesign.org/doc/w3c/xml11/>>.
  - Unicode Consortium. “Unicode Home Page.” En ligne <<http://unicode.org/>>.
  - W3C. Spécification de XML 1.0 (5<sup>e</sup> édition, datée du 26 novembre 2008). En ligne <<http://www.w3.org/TR/REC-xml>>.
  - W3C. Spécification de XML 1.1 (2<sup>e</sup> édition, datée du 29 septembre 2006). En ligne <<http://www.w3.org/TR/xml11>>.
  - W3C. Spécification des schémas XML du W3C (2<sup>e</sup> édition, datée du 28 octobre 2004). En ligne <<http://www.w3.org/TR/xmlschema-0/>>.
  - W3C. “Metadata and Resource Description.” En ligne <<http://www.w3.org/Metadata/>>.
  - W3Schools. “W3Schools Online Tutorials.” Nombreux tutoriels gratuits sur plusieurs normes, du W3C et autres. En ligne <<http://www.w3schools.com/>>.
  - XML.org . “Applying XML and Web Services Standards in Industry.” Portail industriel XML. En ligne <<http://www.xml.org/>>.
  - Yergeau, F.; et al. *Terminologie bilingue XML*. En ligne <<http://pages.videotron.com/fyergeau/w3c/xml10/termino.html>>.
  - [Répertoire d’outils XML gratuits de Lars Marius Garshol \(Norvège\)](#)
- 

